AD-A177 452 THE APPLICABILITY OF MINIMUM-RELATIVE-ENTROPY SPECTRAL 1/1
ESTIMATION: AN ANALYSIS AND CRITIQUE(U) NAVAL RESEARCH
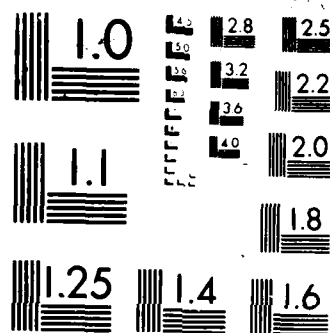LAB WASHINGTON DC N M BLACHMAN 21 JAN 87 NRL-MR-5920

UNCLASSIFIED F/G 20/13 NL

1.0		2.8	2.5
1.1		3.2	2.2
		3.6	2.0
		4.0	1.8
1.25	1.4	1.6	

MICROCOPY RESOLUTION TEST CHART

The Applicability of
Minimum-Relative-Entropy Spectral Estimation:
An Analysis and Critique

NELSON M. BLACHMAN

Computer Science and Systems Branch
Information Technology Branch

DTIC
ELECTE
MAR 1 0 1987
S
D
D.

87 3   9 021

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
NRL Memorandum Report 5920	

6a NAME OF PERFORMING ORGANIZATION	6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION
Naval Research Laboratory	Code 5591	

6c. ADDRESS (City, State, and ZIP Code)	7b. ADDRESS (City, State, and ZIP Code)
Washington, DC 20375-5000	

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL - (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
Office of Naval Research		

8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO.
Arlington, VA 22217	61153N		RR021-05-42	DN980-167

11 TITLE (Include Security Classification)

The Applicability of Minimum-Relative-Entropy Spectral Estimation: An Analysis and Critique

12. PERSONAL AUTHOR(S)
Blachman, Nelson M.

13a. TYPE OF REPORT	13b TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM 10/1/85 TO 11/3/86	1987 January 21	17

16 SUPPLEMENTARY NOTATION

17	COSATI CODES		18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Spectral estimation Signal detection and estimation
			Spectral analysis Signal processing
			Maximum-entropy methods Relative entropy

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The principle of minimizing the relative entropy of the posterior spectral estimate with respect to the prior has been given a good theoretical foundation for stationary gaussian random processes. It is not as firm as one might like however, and its real value thus lies in how accurately it can determine spectra and how much computing is involved. The question remains open as to which applications are suitable for relative-entropy methods, which are not, and why. It can best be answered by trying out this technique along with others on a wide variety of waveforms. Multiple-signal minimum-relative-entropy spectral estimation (MRESE) has been applied to the sum of a signal plus noise to obtain improved estimates of the spectra of both, and provision has been made for weighting their prior estimates. The limit is here investigated as the weight assigned to the prior signal-spectrum estimate approaches zero while the total signal power is required to have a given value. In this limit the prior signal spectrum is found to have no effect on the posterior noise spectrum or signal spectrum. The latter becomes a line at the frequency where the difference between the reciprocals of the posterior and prior estimates of the noise spectral density is least. In addition, a method is found for estimating unknown

(Continues)

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED

22a NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE (Include Area Code)	22c OFFICE SYMBOL
David Mutchler	(202) 767-3490	Code 5591

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted   SECURITY CLASSIFICATION OF THIS PAGE
All other editions are obsolete

☆U.S. Government Printing Office: 1985—507-047

Cont'd

, Over

19. ABSTRACT (Continued)

scaling of the prior spectral estimates. Some aspects of MRESE still in need of investigation are pointed out, and possible approaches are suggested for several of them. *(Keywords:*)

ii

CONTENTS

The Applicability of Minimum-Relative-Entropy Spectral Estimation: An Analysis and Critique

1. Introduction

Minimum-relative-entropy spectral estimation (MRESE) [1] is based on a set of (approximate) values for the autocorrelation function (ACF) calculated from a finite sequence of observed values of a stationary random process, the power spectral density being the Fourier transform of the ACF according to the Wiener-Khinchin theorem [2]. MRESE assumes that the ACF values for longer lags are such as to minimize the relative entropy of the posterior ("final") spectral estimate (i.e., the estimate that takes the foregoing ACF values into account) with respect to the prior ("initial") estimate. This approach to supplying the additional information needed for determining the spectrum is justified by a set of very appealing axioms [1] concerning the effect of new information in determining a posterior spectral estimate from a prior spectral estimate.

Since entropy is the measure of uncertainty, its maximization ensures that as little as possible is assumed about the processes beyond what is known. In the case of a sinusoidal signal in gaussian noise, the signal is generally not gaussian, but the minimization of (6) is nonetheless desirable because of its ease, and, if the amplitude of the signal is unknown, the latter might reasonably be regarded as a sinusoidal gaussian random process. Regardless of the theoretical justifications for the use of MRESE, however, its value lies in the improved spectral resolution that it offers [5] in suitable applications.

The relative entropy of a (posterior or final) probability density function $q(\mathbf{x})$ with respect to a (prior or initial) probability density function $p(\mathbf{x})$ is

$$H(q,p) = \int q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x})}\, d\mathbf{x},$$

where $\mathbf{x}$ can be a vector whose T/τ components represent the values of a signal plus noise at instants spaced uniformly by τ over an interval of length T. We shall take all logarithms to the base e. Thus, the relative entropy (also called the cross entropy, directed divergence, discrimination information, and Kullback-Leibler number) per sample is $\tau H(q,p)/T$ natural units (nits), a nit being 1.4427 bits (binary units). Since $H(q,p)$ is invariant under any invertible change of coordinates [3], it will have the same value if, instead of with the components of $\mathbf{x}$, we deal with the Fourier coefficients of the sequence of values or with the powers and the phase angles of the Fourier components. Since the sampling rate is $1/\tau$, the frequencies of these Fourier components extend from 0 up to the Nyquist frequency $1/2\tau$; any higher-frequency components of the waveform would be aliased into this band by the sampling process.

For a gaussian random process, which, for any given spectrum, has the greatest entropy rate, these phase angles are independently uniformly distributed. Their probability densities in the numerator and denominator of $q(\mathbf{x})/p(\mathbf{x})$ therefore cancel out and can be ignored. The powers y of the Fourier components are independently exponentially distributed with means given by the bandwidth $1/T$ times the power spectral density, say $Q(f)$ or $P(f)$, respectively. Substituting the probability density functions $p(y) = \dfrac{T}{P(f)}\, e^{-Ty/P(f)}$ and $q(y) = \dfrac{T}{Q(f)}\, e^{-Ty/Q(f)}$ for $y \geq 0$ and 0 otherwise for the power of the Fourier component of frequency f into $\log \dfrac{q(y)}{p(y)}$ and averaging over the posterior distribution $q(y)$, we find that these components contribute

$$\mathop{\mathbf{E}}_{q(y)}\left\{ \frac{Ty}{P(f)} - \frac{Ty}{Q(f)} - \log\frac{Q(f)}{P(f)} \right\} = \frac{Q(f)}{P(f)} - \log\frac{Q(f)}{P(f)} - 1$$

to the relative entropy of the Fourier component at frequency f. Note that $u - \log u - 1$ is a convex function of u for $u > 0$ with its minimum at $u = 1$. In that neighborhood it is approximately $(u - 1)^2/2$, and it rises to ∞ as $u \to 0$ and as $u \to \infty$, thus ensuring that $Q(f) > 0$.

Approximating the sum over all frequencies (spaced by $1/T$) by an integral, we find that the per-sample relative-entropy rate is

$$\tau \int\limits_{0}^{1/2\tau} \left[\frac{Q(f)}{P(f)} - \log\frac{Q(f)}{P(f)} - 1 \right] df \; ,$$

which is τ times the Itakura-Saito distortion [4]. The factor τ here cancels the df dimensionally, causing this quantity to be measured in nits; it has also been described as the *normalized* Itakura-Saito distortion. Hence, for gaussian random processes, minimization of the relative entropy means minimization of the Itakura-Saito distortion.

Section 2 discusses the extension of MRESE to the case where one has not only values of the ACF of the sum of two waveforms—signal and noise—but also prior estimates of their separate spectra as well as weights for the latter, and one wants posterior estimates for their separate spectra. Section 3 studies the limit as the weight for the prior estimate of the spectrum of a sinusoidal signal of unknown frequency is allowed to approach zero because the broad prior estimate, reflecting ignorance as to its frequency, is obviously a very poor approximation to the spectrum of a sinusoid. Section 4 deals with the resulting posterior estimate of the signal's spectrum, and Sec. 5 with computation of estimates of the signal's frequency and strength. Section 6 discusses the inconsistency between the single-signal MRESE estimate of the spectrum of the sum of the two waveforms (signal and noise) and the sum of their separate estimates, and Sec. 7 determines the appropriate scale factors to be used when the prior estimates of the signal and noise spectra are subject to unknown scaling, as in the case where there may be an unknown but constant attenuation in each transmission path. Section 8 discusses various other problems yet to be resolved in regard to the application of MRESE, and finally Sec. 9 summarizes these problems and the progress that has been made in the preceding sections.

2. Multisignal MRESE

Multisignal minimum-relative-entropy spectral estimation [6] uses $M + 1$ values $R(r\tau)$ (with $r = 1, 2, \cdots, M$) of the autocorrelation function of the observed signal-

plus-noise process along with prior estimates $P_s(f)$ and $P_n(f)$ of the signal and noise spectra, respectively, and [7] positive weights w_s and w_n (possibly depending on the frequency f) for these prior estimates—or, more precisely, for the Itakura-Saito distortions

$$D(Q_s, P_s) = \tau \int_0^{1/2\tau} \left[\frac{Q_s(f)}{P_s(f)} - \log \frac{Q_s(f)}{P_s(f)} - 1 \right] df \tag{1}$$

and

$$D(Q_n, P_n) = \tau \int_0^{1/2\tau} \left[\frac{Q_n(f)}{P_n(f)} - \log \frac{Q_n(f)}{P_n(f)} - 1 \right] df \tag{2}$$

that they suffer when replaced by $Q_s(f)$ and $Q_n(f)$—to obtain the posterior spectral estimates

$$Q_s(f) = \frac{1}{\dfrac{1}{P_s(f)} + \dfrac{1}{w_s} \sum_{r=0}^M \beta_r \cos 2r\pi f \tau} \tag{3}$$

and

$$Q_n(f) = \frac{1}{\dfrac{1}{P_n(f)} + \dfrac{1}{w_n} \sum_{r=0}^M \beta_r \cos 2r\pi f \tau}, \tag{4}$$

where the $\{\beta_r\}$ (Lagrange multipliers used in the distortion minimization) are chosen so that

$$\int_0^{1/2\tau} [Q_s(f) + Q_n(f)] \cos 2r\pi f \tau df = R(r\tau) \tag{5}$$

for $r = 0, 1, \cdots, M$; where τ is the sampling interval. Because the second variation of the weighted Itakura-Saito distortion

$$\tau \int_0^{1/2\tau} w_s(f) \left[\frac{Q_s(f)}{P_s(f)} - \log \frac{Q_s(f)}{P_s(f)} - 1 \right] df + \tau \int_0^{1/2\tau} w_n(f) \left[\frac{Q_n(f)}{P_n(f)} - \log \frac{Q_n(f)}{P_n(f)} - 1 \right] df \tag{6}$$

with respect to $Q_s(f)$, viz., $\tau w_s / Q_s^2(f)$, is positive while the constraints (5) are linear in $Q_s(f)$, a posterior spectrum of the form (3) obtained by setting the first variation equal to zero will not only minimize (1) but will yield the unique minimum. Similarly, a posterior spectrum of the form (4) uniquely minimizes (2).

The $M+1$ conditions (5) on $Q_s(f) + Q_n(f)$ imply that

$$Q_s(f) + Q_n(f) = 2\tau R(0) + 4\tau \sum_{r=1}^\infty R(r\tau) \cos 2r\pi f \tau$$

with $R(0), \cdots, R(M\tau)$ having the measured values and $R((M+1)\tau), R((M+2)\tau), \cdots$ having the values that minimize (6). Thus, the known values of the autocorrelation function determine the gross form of the posterior total spectrum while the fine structure is implied by the minimization of the weighted relative entropy.

3. Small Weight for Prior Signal Spectrum

In some applications, such as HF radar, the signal may often be well approximated by a sinusoid of unknown strength and (Doppler) frequency. In such a case, the limit of

(3) as $w_s \to 0$ is of considerable interest because it has been observed [7] that, when w_s is small, $Q_s(f)$ tends to be narrow and thus provides a sharp indication of the signal frequency f_s. When f_s is unknown and $P_s(f)$ is therefore taken to be flat across the band of possible values, it should be appropriate to assign only a small weight to this clearly inaccurate broad estimate of a spectrum that is known to be narrow.

We cannot simply set $w_s = 0$, however, as there would then be no constraint on the normalized Itakura-Saito distortion (1) and so $Q_s(f)$ would no longer have to remain positive. Another reason for a careful approach to the limit is that the forms (3) and (4) along with the constraints (5) do not uniquely determine the Lagrange multipliers $\{\beta_r\}$ despite the convexity of (1) and (2) and the consequent uniqueness of the minimum of (6). More than one set of values for the $\{\beta_r\}$ in (3) and (4) will satisfy (5) because of the nonlinearity of (3) and (4) as functions of the $\{\beta_r\}$. In particular, as $w_s \to 0$, there are generally two types of solutions—one with values for the $\{\beta_r\}$ that likewise approach 0 and the other with values for the $\{\beta_r\}$ that do not. In the latter case, $Q_s(f)$ generally approaches 0 for every f, and the constraints (5) are satisfied by (4) alone. This solution, however, yields a positive weighted Itakura-Saito distortion (6) while the other, with $\beta_r \to 0$, may make $Q_n(f) = P_n(f)$ and may then satisfy (5) by giving (3) a suitable form, thus making (6) zero. [However, there will be no solution of the latter type if there is no nonnegative $Q_s(f)$ of the form (3) that, when added to $Q_n(f) = P_n(f)$, satisfies (5); e.g., when $P_n(f)$ is too large.]

Neither of these two types of spectral estimates is satisfactory, as one yields a vanishing posterior signal-spectrum estimate, and the other fails to modify the noise-spectrum estimate on the basis of the information (5). Accordingly, before letting $w_s \to 0$, we impose the constraint

$$\int_0^{1/2\tau} Q_s(f)\, df = S \tag{7}$$

on $Q_s(f)$ (with the help of the Lagrange multiplier λ) in addition to the constraints (5) so that the total estimated signal power will be S and cannot go to zero with w_s even though we restrict our attention to posterior spectra of the form (3) and (4) with the $\{\beta_r\}$ not all going to zero so that $Q_n(f)$ will be influenced by the ACF values (5). Being another linear constraint like (5), (7) ensures that the minimization of the weighted Itakura-Saito distortion (6) still yields a unique set of posterior spectral estimates.

The effect of (7) is to add to the β_0 in (3) but not to that in (4) a constant λ which is to be given the value satisfying (7). [In the case of a flat $w_s(f)$, a flat $1/P_s(f)$ can be absorbed into λ.] λ must be large enough to ensure that the denominator of (3) never becomes negative—but only barely large enough when w_s is small, for otherwise (3) will approach zero at all frequencies and will not satisfy (7). When $w_s(f)$ is small, $Q_s(f)$ then differs significantly from zero only near the frequency, say f_0, where the summation in its denominator is small, i.e., near the minimum of the summation in (4). Hence, $Q_s(f)$ automatically becomes a line spectrum with power S at frequency f_0, which is thus the minimum-relative-entropy estimate of the signal's frequency. As $w_s(f)$ goes to zero at all frequencies, its effect and that of $P_s(f)$ on $Q_s(f)$ vanish, and it becomes unnecessary to choose a prior estimate or weighting function for the signal spectrum.

Since, with (7) included, the $\{\beta_r\}$ do not approach zero, $Q_n(f)$ is able to reflect the information contained in (5). $Q_s(f)$ too is able to reflect this information because (3) then differs from zero only where the quantity multiplying $1/w_s$ in its denominator is small (on account of the inclusion of λ). Thus, the resulting posterior spectral estimates

have the desirable aspects of both the solution for which $\beta_r \rightarrow 0$ and that for which it does not, and $Q_s(f)$ has the right form for a sinusoidal signal.

4. Posterior Signal Spectrum

When w_s is small but is still positive, the principal denominator of (3) has a global minimum at, say, f_0 that does not quite reach zero, and in that neighborhood it is approximately a quadratic function of the frequency f. Hence, $Q_s(f)$ has the shape of a witch of Agnesi [like the Cauchy density function $p(x) = \pi^{-1}/(1 + x^2)$] centered at f_0. The height of the witch is proportional to S^2/w_s, and its width is proportional to w_s/S when w_s is small. As long as S is small, the apportionment of this amount of power to the posterior signal spectrum $Q_s(f)$ will hardly affect the posterior noise spectrum $Q_n(f)$, and the latter can be determined in the single-signal manner [1] just as if no signal were present, i.e., as (4) with the $\{\beta_r\}$ determined by (5) with $Q_s(f) = 0$.

When S is significant in comparison with the total noise power. its effect on the $\{\beta_r\}$ must be taken into account, as it will contribute $S \cos 2r\,\pi f_0\tau$ to (5). In any case, the location f_0 of the global minimum of

$$w_n(f)\left\{\frac{1}{Q_n(f)} - \frac{1}{P_n(f)}\right\} \tag{8}$$

[which is the summation in (3)] is the MRESE estimate of the frequency f_s of the signal. Without the condition (7) (and without λ) (8) might remain positive, making $Q_s(f) = 0$ at all frequencies, or it might go to zero at some frequency f_0, giving the posterior signal power at that frequency the indeterminate value $0/0$. Thus, (7) resolves any such indeterminacy.

5. Estimation of Signal Strength and Frequency

It remains to find a suitable way to choose S when the signal strength is unknown (and may be 0) and a way to assign an accuracy to the estimate f_0 of the signal's frequency f_s as well as to provide a measure of the reliability of the determination that a signal is present or absent. A possible way to estimate the signal power S would be to determine the S that minimizes the weighted Itakura-Saito distortion (2). Experimental testing may show whether this approach is useful. If the resulting value of S exceeds some threshold, it can be deemed to indicate the presence of a signal, and otherwise its absence.

The second derivative of (6) with respect to S at its minimum (with $w_s = 0$) may be inversely proportional to the variance of this estimate of S, since, for a normal distribution, the relative entropy resulting from a shift equals half the ratio of the square of that shift to the variance. More generally, the second derivative at zero shift of the relative entropy of a distribution with respect to a shifted version, which is the Fisher information of that distribution [8, p. 1010], cannot, by the Cramér-Rao inequality [8, p. 943], be less than the reciprocal of the variance, with equality only in the normal case, and so this second derivative provides a lower bound for the variance.

The determination of the signal-frequency estimate f_0 is straightforward when S is small, but an iterative procedure appears necessary for larger S. For this purpose the f_0 for small S can be used to subtract $S \cos 2r\,\pi f_0\tau$ from the initial values of $R(r\,\tau)$. A new f_0 can then be obtained from the minimum of the resulting function (8), and the process can be continued for an increasing sequence of values of S. This procedure will

5

yield a unique estimate of the signal's frequency even though it might otherwise not be unique. The iterative estimation of f_s as the total signal power S is gradually increased from zero may well require less computing time than the search [9] for the spectra (3) and (4) that satisfy (5), as it has the advantage of not needing a prior signal spectrum nor relative weights for prior signal and noise spectra. It remains to be seen how well this approach works and in what sort of applications it performs best. (See also [10].)

The iterative process for determining the estimate f_0 of the signal frequency f_s and the estimate $Q_n(f)$ of the noise spectrum begins by computing the initial estimate $Q_n^{(0)}(f)$ from $P_n(f)$ and the observed values of $R(r\tau)$ via the Levinson technique utilized in the second algorithm of [9] and then determining the initial estimate $f_0^{(1)}$ as that f for which $w_n(f)[1/Q_n^{(0)}(f) - 1/P_n(f)]$ is least. At the ith succeeding stage of the iteration, the observed values of $R(r\tau)$ are reduced by $S^{(i)}\cos 2r\pi f_0^{(i-1)}\tau$, $Q_n^{(i)}(f)$ is computed from them and $P_n(f)$, and $f_0^{(i)}$ is determined as the f that minimizes $w_n(f)[1/Q_n^{(i)}(f) - 1/P_n(f)]$. The iteration is continued for a sequence of values $S^{(i)}$ of S increasing from $S^{(0)} = 0$ until $S^{(i)}$ reaches the desired total signal power S, with decreasing increments as this value is approached.

6. Consistency of Multiple-Signal and Single-Signal Spectral Estimates

One might expect the sum of the spectra (3) and (4) to equal the estimate

$$Q(f) = \frac{1}{\dfrac{1}{P_s(f) + P_n(f)} + \sum_{r=0}^{M} \beta_r' \cos 2r\pi f\tau} \tag{9}$$

of the spectrum of the total observed process subject to the same constraints

$$\int_0^{1/2\tau} Q(f)\cos 2r\pi f\tau\, df = R(r\tau). \tag{10}$$

A simple example, $M = 1$, suffices to show, however, that, in the case of flat prior spectral estimates and constant weights, for example, (9) is in general not the sum of (3) and (4). When (7) is very small, the posterior signal spectrum has very little effect on the total posterior spectrum, and there is almost no difference between (9) and the sum of (3) and (4). When w_s is small but S is not, however, $Q_s(f)$ includes a substantial spectral line while (9) does not, and so there is again a significant difference between the sum of the multisignal minimum-relative-entropy estimates and the single-signal estimate. The knowledge that a spectral line may be present fundamentally alters the nature of the estimate.

Another approach [11] to multiple-signal relative-entropy spectral estimation, however, begins with the assumption that the sum of the posterior spectra and cross-spectra is the MRESE of the sum of the signal and noise based on the prior estimate of the spectrum of the sum. Musicus and Johnson divide the difference between the total posterior and total prior spectra among the posterior signal and noise spectra and cross-spectra by what is, in effect, Wiener filtering. The resulting posterior signal and noise spectra therefore differ from (3) and (4), and cross-spectra arise even in the absence of prior cross-correlation. This approach does not incorporate weighting, and so it is not possible to remove the influence of the prior signal-spectrum estimate upon its posterior spectral estimates.

7. Unknown Scales for Prior Spectra

In some applications, such as the processing of noisy speech, the signal and noise will suffer unknown constant power gains $10 \log_{10} \gamma_s$ and $10 \log_{10} \gamma_n$ decibels. Here it may be appropriate to use $\gamma_s P_s(f)$ and $\gamma_n P_n(f)$ as the prior spectral estimates instead of $P_s(f)$ and $P_n(f)$. Substituting these into (1), (2), and (6) and minimizing (6) under the constraints (5) with respect to $Q_s(f)$ and $Q_n(f)$, we get (3) and (4) with these substitutions for $P_s(f)$ and $P_n(f)$. Setting the derivatives of (6) with respect to γ_s and γ_n equal to zero, we find

$$\gamma_i = \frac{\displaystyle\int_0^{1/2\tau} w_i(f) \frac{Q_i(f)}{P_i(f)} df}{\displaystyle\int_0^{1/2\tau} w_i(f)\, df} \tag{11}$$

for $i = s$ and n. Hence,

$$Q_i(f) = \frac{1}{\dfrac{\displaystyle\int_0^{1/2\tau} w_i(f')\, df'}{\displaystyle\int_0^{1/2\tau} w_i(f')Q_i(f')\, df'\, /P_i(f')} + \dfrac{\displaystyle\sum_0^M \beta_r \cos 2r\pi f\, \tau}{w_s(f)}} \tag{12}$$

for $i = s$ and n. Since the second partial derivative of (6) with respect to $1/\gamma_i$ is $\tau\gamma_i^2 \int_0^{1/2\tau} w_i(f)\, df > 0$, we see not only that (11) yields a minimum for (6) but also that this minimum is unique.

If either prior spectral density $P_i(f)$ is independent of frequency, the effect of (11) is simply to replace it in (3) or (4) with the value of the corresponding spectral estimate $Q_i(f)$ averaged over the frequency band from 0 to $1/2\tau$ with the weighting $w_i(f)$. In the case of noise alone with a flat $w_n(f)$, this average value of $Q_n(f)$ is just $2\tau R(0)$. Otherwise it is necessary to search for the values of γ_s and γ_n that are consistent with (12) as well as for the values of the $\{\beta_r\}$ that cause the sum of (12) for $i = s$ and for $i = n$ to satisfy (5).

To simplify the computation one might instead have chosen γ_s and γ_n to normalize the prior spectral estimates, i.e., to give them each a unit total power. This procedure, however, would increase (6); and the choice of a *unit* total power would introduce an inadmissible arbitrary effect on the result.

8. Problems Remaining to be Resolved

8.1. *Inexactness of Autocorrelation-Function Values*

Although MRESE appears to be based on a solid logical foundation [1], it must be recognized that ACF values computed from a finite record will not exactly equal the true values of the autocorrelation function. Thus, not only are ACF values for lags longer than the available record unknown, but those for shorter lags (especially for only slightly shorter lags) are known imprecisely, contradicting an assumption of the MRESE approach. Efforts [12],[13] have been made to take into account the inexactitude of ACF

values by imposing an upper bound on a quadratic function of their departures from the initial estimates and then determining the set of ACF values which, subject to this condition, minimizes the relative entropy of the posterior spectral estimate with respect to the prior.

8.2. *Logical Assignment of Weights to Prior Spectral Estimates*

While the procedure just described produces a unique posterior spectral estimate , it supposes that the corrections of the approximate ACF values should combine to minimize the relative entropy. It is equally likely. however, that they will combine to maximize it; but it is much *more* likely that they will move the ACF vector (the set of ACF values) in a nearly orthogonal direction that hardly affects the relative entropy but may have significant effects upon the shape of the posterior spectral estimate. Thus. a better idea of the implications of the approximate nature of the ACF values should be obtainable by exploring the ACF-vector space in the neighborhood of the estimated point, and attaching to each spectral value an uncertainty given by the variety of resulting posterior spectral values.

Such a procedure, however, would involve a far greater amount of computation than the more naive MRESE approaches, but it could provide information as to the accuracy of the posterior spectral estimate and could thus yield a posterior weighting function that might enable the posterior spectral estimate to be used as a prior estimate with new observations in the same way as the previous prior estimate [7]. It may, on the other hand, be necessary to use each of a large number of previous posterior spectra as the new prior in order to obtain a reliable picture of the new posterior spectra implied by them.

Apart from this approach (which ignores correlations between errors in the values of the ACF for different lags) and that of Secs. 3 and 4 with $w_s \rightarrow 0$, a logical basis is needed for assigning weights to the prior spectral estimates for the signal and the noise just as the reciprocals of the variances of the residuals serve as weights for overdetermined linear equations. It remains to be seen whether weighting inversely proportional to the variance of the prior estimate at each frequency (but still ignoring covariances) might yield satisfactory posterior spectral estimates.

8.3. *Criterion for Admissibility for Prior Spectral Estimates*

As formulated, MRESE has imposed no requirements concerning the accuracy or reliability of the prior spectral estimates that it uses, and so there is as yet no logical reason to prefer one prior estimate (such as the maximum-entropy flat estimate) over another. To avoid this arbitrariness and to ensure the usefulness of the resulting posterior estimates, it is necessary that the prior estimates have adequate credentials. Suitably assigned and utilized weights might provide a basis for the admissibility of prior estimates. In fact, it might turn out that the classical spectral estimate [14] based on the same ACF data could serve well as a prior for obtaining a much sharper posterior minimum-relative-entropy estimate.

8.4. *The Relative Nature of Weighting*

By introducing the weighting functions $w_s(f)$ and $w_n(f)$ for the Itakura-Saito distortions (1) and (2), we are able to place greater or lesser reliance on $P_s(f)$ or $P_n(f)$ at each frequency. These weights are only relative, however, as multiplication of both by the same constant will leave the posterior spectral estimates unchanged: there is as yet no way to introduce absolute measures of the credibilities of the prior estimates. An absolute standard with respect to which weights might somehow be expressed is the flat

prior estimate that underlies maximum-entropy spectral estimation. It might be appropriate for the posterior spectral estimate based on such a prior to be independent of any weighting, as the constant spectral density already represents maximum entropy—maximum uncertainty as to the spectrum. Although (3) and (4) depend on $w_s(f)$ and $w_n(f)$ even in this case, a new approach to weighting might remove that dependence.

8.5. *Other Methods of Spectral Estimation*

Many other forms of spectral estimation besides MRESE and classical methods [14] have been devised [15]-[19]—each with its own devotees, who find it well suited to particular applications. When a spectrum can be assumed to be described by a finite number of unknown parameters, its estimation can be treated as a parameter-estimation problem. MRESE performs just such a parameter fitting and does it very well for autoregressive spectra of the form (3) and (4), to which it leads, especially when the order M of the autoregressive process is known. A new approach [20] to maximum-entropy spectral analysis offers confidence intervals that may help to resolve some of the weighting problems mentioned above. For estimating other kinds of spectra, on the other hand, another method may turn out to be superior to MRESE. It needs to be determined experimentally why and in what cases MRESE does well.

8.6. *Iterative Use of MRESE*

In OTH-radar detection MRESE has been found misleading at low signal-to-clutter ratios if the posterior estimates of the signal and clutter spectra (3) and (4) are used as new priors [21], since a weak signal can then show up more and more in the posterior clutter spectrum rather than in the posterior signal spectrum, thus escaping detection. To avoid this outcome, it seems preferable not to iterate but, instead, to use the average of all of the autocorrelation values for each lag—or to average all of the posterior spectra obtained without iterating. Experience should show when such an approach is needed and when iteration is satisfactory.

8.7. *Limitation on the Order M of the Autoregressive Spectrum*

In applications it has been found necessary to choose the right order M for the autoregressive filter whose frequency response represents the minimum-relative-entropy or the maximum-entropy estimate of the spectrum of interest, as too large an M yields too many spectral peaks and valleys—features that exhibit no stability with repeated analysis or increase in M. MRESE therefore makes use of only a limited number of values $R(0), \cdots, R(M\tau)$ of the ACF. The information contained in the estimates of the ACF for longer lags is ignored, and it seems desirable to find a way, within the MRESE framework, to utilize this additional information.

The spectrum that fits these $M + 1$ ACF values will have autocorrelation-function values for longer lags that do not quite match the available ACF estimates $R((M+1)\tau)$, $R((M+2)\tau)$, $\cdots$ of longer lags, and it may therefore be useful to adjust $\beta_0, \cdots, \beta_M$ to reduce the mismatch while introducing not too large discrepancies for $R(0), \cdots, R(M\tau)$. The result, however, may no longer be describable as MRESE but, rather, as curve fitting unless an information-theoretical basis can be devised for the foregoing parameter adjustment. An alternative approach would be to use an average of posterior spectra based on $R(k\tau)$, $R((k+1)\tau)$, $\cdots$, $R((k+M)\tau)$ for $k = 0, 1, \cdots$. Again the value of each approach will have to be determined by its success in applications.

9

8.8. *Disregard of Available Information*

MRESE has been described as making full use of all available ACF information. unlike classical spectral analysis, which windows the ACF to make it fall smoothly to zero at lags exceeding the length T of the available record, and as, in effect, determining values for the ACF of longer lags in a manner that minimizes the amount of information they imply concerning the spectrum. We see, however, that MRESE throws away even more information than classical spectral analysis, viz., $R(r\tau)$ for $r > M$, where $M\tau$ is generally small compared with T, and it instead estimates these ACF values. It does make very good use of $R(0), \cdots, R(M\tau)$ and is thus able to provide better spectra than classical methods in suitable applications. Some of the foregoing approaches may. on the other hand, be able to do even better by making use of the ACF values that are redundant for autoregressive processes of order M but otherwise contain useful information.

8.9. *Oversampling*

The problem of having too many ACF values (if it is indeed a problem) can be exacerbated by reducing the interval τ between the samples of the record of a waveform of duration T. While the most convenient sampling interval is Nyquist's when the waveform can be assumed to be bandlimited, the spectral density may fall only slowly toward zero with increasing frequency, thus raising a question as to what should be considered its cut-off frequency and what should accordingly be the sampling interval τ. A shorter sampling interval ought to yield at least as much information about the waveform and its spectrum as a longer one, but some modification of MRESE may be necessary in order to take advantage of such an increase in the amount of information available *without thereby introducing false and unstable spectral peaks and valleys.*

9. Summary

Despite its success in some applications, we thus see that MRESE might benefit from the investigation of means for taking care of:

- The approximate nature of the available values of the autocorrelation function,

- Use of the information contained in the ACF of lags beyond those now utilized,

- The superabundance of information due to a shorter sampling interval τ,

- Posterior spectra of the form (3) and (4) satisfying (5) but not minimizing (6),

- Nongaussian waveforms,

- Specification of credentials or qualifications to be required of prior estimates,

- Absolute rather than the present relative weighting of prior spectral estimates,

- Logical assignment of weights to prior spectral estimates, and

- Determination of weights for the posterior spectra when used as priors.

The first five problems listed above are common to both MRESE and maximum-entropy spectral estimation. Hence, any solution of them would benefit both. Some possible approaches to the solution of several of the foregoing problems have been presented in the preceding sections. In Secs. 3 and 4 the problem of assigning a prior spectral estimate for a sinusoidal signal has been solved by investigation of the limit as the weight given to that estimate is allowed to approach zero. It is then unnecessary to introduce either a prior spectral estimate for the signal or a weight for it. In Sec. 7 we determined the best choice of gain factors for prior spectral estimates when the latter are subject to unknown scaling. All of these ideas need to be tested experimentally in a wide variety of applications along with the standard form of MRESE and other methods [14]-[19] of spectral analysis to compare them as to resolution, reliability, and computational requirements.

Acknowledgments

References

[1] J.E. Shore and R.W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 26-37 (1980). Cf. comments and corrections, *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 942-943 (1983).

[2] N.M. Blachman, *Noise and Its Effect on Communication*, chap. 3. New York: McGraw-Hill Book Co., 1966; 2nd ed., Malabar, Fla.: Krieger Publishing Co., 1982.

[3] N.M. Blachman, "The amount of information that y gives about X," *IEEE Transactions on Information Theory*, vol. IT-14, pp. 27-31 (1968).

[4] F. Itakura and S. Saito, "An analysis-synthesis telephony based on maximum-likelihood method," *Rept. Sixth Int. Cong. Acoust.*, 1968, vol. 2, sect. C-5-5, p. 17.

[5] J.E. Shore, "Minimum cross-entropy spectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 230-237 (1981).

[6] R.W. Johnson and J.E. Shore, "Minimum cross-entropy spectral analysis of multiple signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 574-582 (1983).

[7] R.W. Johnson, J.E. Shore, and J.P. Burg, "Multisignal minimum-cross-entropy spectrum analysis with weighted initial estimates," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 531-539 (1984).

END

4-87

DTIC